Zooming in on the O–O Bond Formation—An Ab Initio Molecular Dynamics Study Applying Enhanced Sampling Techniques

Schilling, Mauro ; Cunha, Richard A ; Luber, Sandra

Abstract: Mastering artificial water oxidation is a key step on moving away from fossil fuels toward a carbon emission-free society. Unfortunately, the crucial chemical transformation of this reaction, the O–O bond formation, is still not well understood, even though there are various known active water oxidation catalysts, such as Ru-based catalysts bearing a Py5 ligand. Those were recently investigated both experimentally and using a static density functional theory (DFT) approach based on geometry optimizations. In this work, we shed light on the O–O formation catalyzed by those Ru-based complexes, utilizing enhanced sampling techniques such as the Bluemoon ensemble and metadynamics together with high-performance DFT-based molecular dynamics simulations. This allowed unprecedented detailed insights into the process of the oxygen–oxygen bond formation and also extended the view on the reaction network and the flexibility of the product state because of the consideration of the dynamics at ambient conditions. Our model system contained both the catalyst and a large number of explicit water molecules which can participate in the reaction and stabilize intermediates. Moreover, it is demonstrated how crucial the choice of the collective variable is in order to capture relevant features of the studied reaction.

DOI: https://doi.org/10.1021/acs.jctc.9b01207

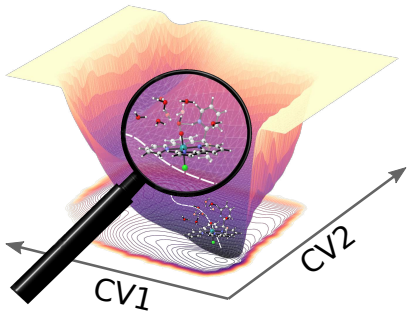
Zooming in on the O−O Bond Formation - an ab initio Molecular Dynamics Study Applying Enhanced Sampling Techniques

Mauro Schilling , Richard A. Cunha, and Sandra Luber*

University of Zurich, Department of Chemistry, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

E-mail: sandra.luber@chem.uzh.ch

Table of Content



Abstract

Mastering artificial water oxidation is a key step on moving away from fossil fuels towards a carbon emission free society. Unfortunately, the crucial chemical transformation of this reaction, the O−O bond formation, is still not well understood, even though there are various

known active water oxidation catalysts, such as Ru-based catalysts bearing a Py5 ligand. Those were recently investigated both experimentally and using a static density functional theory (DFT) approach based on geometry optimizations. In this work we shed light on the O−O formation catalyzed by those Ru-based complexes, utilizing enhanced sampling techniques such as the *Bluemoon* ensemble and metadynamics together with high-performance DFT-based molecular dynamics simulations. This allowed unprecedented detailed insights into the process of the oxygen-oxygen bond formation and also extended the view on the reaction network and the flexibility of the product state due to the consideration of the dynamics at ambient conditions. This was possible since the model system contained both the catalyst and a large number of explicit water molecules which can participate in the reaction and stabilize intermediates. Moreover, it is demonstrated how crucial the choice of the collective variable is in order to capture relevant features of the studied reaction.

Introduction

The quest for the development of renewable energy sources is among the great challenges of the 21st century. Water splitting thereby promises to be a suitable solution, since it is the fundamental basis of the photosynthesis process, which is nature's primary mechanism to store sun-light energy in the form of chemical bonds. From a chemical point of view, the process might be subdivided into water oxidation and reduction, both of which usually require different conditions e.g. pH value and electron acceptors or electron donors. Those factors make up the need for compartmentalization. For this reason, catalysts have been developed with the aim to either drive water oxidation or reduction.

In the past, we have throughly studied the thermodynamics and kinetics of a series of Ru-based water oxidation catalysts (WOCs).[1,2] Those WOCs feature a pentapyridyl ligand (Py5R = 6,6″–(R-(pyridin–2–yl)methylene)di–2,2′-bipyridine), where R is either a methyl or methoxyl substituent and a chlorido ligand to complete the octahedral coordination of

the metal center (see Figure S1 in the supplementary information). In the later the two catalysts are referred to as **{Ru(Py5OMe)}** and **{Ru(Py5Me)}**. The catalytic performance of the hydrolyzed WOCs, i.e. where the chlorido ligand was replaced by a water molecule, was assessed in a standard chemical water oxidation setup. Despite the innocent nature of the substituent at the Py5R ligand, different catalytic performances (TON/TOF: **{Ru(Py5OMe)}**: 8 / 0.037 s^{-1} and **{Ru(Py5Me)}**: 1 / 0.021 s^{-1})[1] were found. On the other hand, the none hydrolyzed **{Ru(Py5OMe)}** showed a TON/TOF of 24 / 0.710 s^{-1}, while no TON/TOF could be determined for **{Ru(Py5Me)}** under the same conditions. In the case of **{Ru(Py5OMe)}** a striking oxidative efficiency of above 95% was achieved.[1] The unexpected difference in catalytic activity caused by the Py5R sprouted our curiosity. Spectroscopic measurements in combination with our simulations suggested that the decoordination of the pyridine ligand is likely if the chlorido ligand is not hydrolyzed.[1] A dangling base in close proximity to the metal-oxo species is a common structural motif which is thought of to facilitate the O – O bond formation by a water nucleophilic attack (WNA).[3–6] While, like for most WOCs, no catalytic intermediates were isolated the WOCs are still a valuable target to study the effect dangling pyridine and its potential role in the O – O bond formation. The in-depth study of those catalysts allowed for the *in silico* design of novel WOCs by proposing modifications of the dangling base to lower the activation barrier of the WNA while keeping the overall thermodynamics about the same.[2] For a general overview of computational studies focusing on the water oxidation mechanism we refer the interested reader to Refs.[7–12]

In the previous study, catalytic intermediates as well as the corresponding transition states (TSs) were simulated by means of density functional theory (DFT) calculations, whereby the energetic contributions of the solute-solvent interaction were approximated by the conductor-like-screening-model (COSMO).[13,14] Within said method directed solute-solvent interactions such as hydrogen bonding are not explicitly accounted for. This can have profound consequences on the obtained structures. For example, in the structure of the

reactant (R), a Ru^{V}=O species, the lone pair of the pyridyl is aligned with the oxo-ligand, implying weak electrostatic interactions between them. Another limitation of the method became evident while attempting to model the TSs. Localization of them was unsuccessful unless at least one additional water molecule was included in the model system. The inclusion of a limited amount of explicit solvent molecules is a common strategy when it comes to modeling TSs and minimum energy pathways (MEPs) connecting the two states.[15–17] The limitations of this approach have been investigated by Hodel *et al.* in the context of ligand exchange reactions on a cobalt based WOC, where the whole first solvation shell was treated explicitly with a static framework and compared to metadynamics-accelerated DFT-MD simulations.[18] The inclusion of an additional solvent molecule in the static simulation of the TS of the WNA catalyzed by **{Ru(Py5OMe)}** had also affected the structures of the associated reactant (AR) and product (P), the other extrema of interest (see Figure 1). The inability of the protic groups, i.e. the water molecules or the newly formed hydroperoxo ligand, to engage in hydrogen bonding with surrounding solvent resulted in an overly stable hydrogen bonding network that is conserved among the three structures of the extrema (AR, TS, P). While this observation is expected within the applied method, there is no foundation to expect this to be true for the real system. The shortcomings of this approach with respect to the description of the solvent dynamics and conformational diversity of the states of interest therefore need to be addressed.

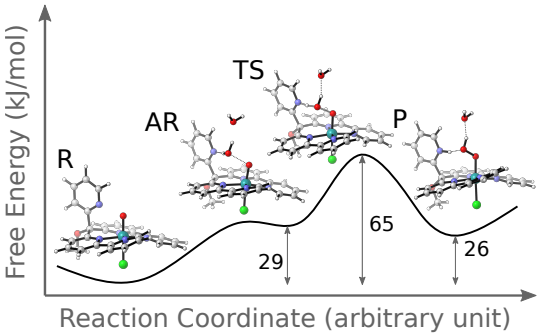


Figure 1: Visualization of the relevant intermediates of the O−O bond formation catalyzed by **{Ru(Py5OMe)}**. The structures are reproduced from geometrical information published by Gil-Sepulcre *et al.* and so are the relative free energies.[1]

In the current study we improve the understanding of the WNA catalyzed by **{Ru(Py5OMe)}** by applying state of the art DFT-based molecular dynamics (DFT-MD), usually called *ab initio* molecular dynamics (AIMD) simulations. Since the O−O bond formation is an activated process, enhanced sampling methods such as the *Bluemoon* ensemble and metadynamics (MetaD) were used. In the following we discuss in-depth how to apply those methods to a complex chemical system in order to elucidate the role of the solvent arrangement, dynamics, and the intramolecular base in the O−O bond formation mechanism. Studies employing AIMD enhanced sampling have been known in the literature for reactions such as the formamide dehydrogenation, the trifluoromethylation of thiols and acetonitrile or the capture of CO_2 by frustrated Lewis pairs.[19–22] However, studies in the context of water oxidation are rare. One of the earliest studies dates back to 2010 and was conducted by Vallés-Pardo *et al.* They used MetaD to simulate an initial reaction path of the O−O bond formation by a WNA which was then refined by *Bluemoon*.[23] An analogues approach was used by Piccinin *et al.* (see Section Metadynamics Simulations).[24] In a recent study Govindarajan *et al.* used *Bluemoon* to model a WNA. In their case, a solvent molecule acted as a proton relay between the nucleophile and a dangling carboxylate, a situation analogous to the WOCs studied in this work.[25] Employing an empirical valence bond MD (EVB-MD)[26] Zhan *et al.* proposed for the same WOC as the one studied by Govindarajan, a nucleophilic attack of a hydroxide on the carboxylate which then transfers one of its oxygen atoms to the metal-oxo species forming the O−O bond.[27] Oxygen-atom transfer reactions such as the one discussed before have recently been suggested as an alternative to the classical WNA.[28] To this extent Govindarajan *et al.* studied the formation of pyridyl-N-oxide i.e the transfer of the metal-oxo ligand to a dangling pyridyl (N_{py}−O) by constrained DFT-MD (*Bluemoon*).[29] This species is then expected to form the O−O bond formation. An analogue mechanism could also be imagined for the Py5-based WOC, but has not been considered so far in previous studies and is beyond the scope of the current study. In 2019, Shao *et al.* modeled the catalytic cycle of a Ru-based WOC, the ligand of which is covalently linked to an or-

ganic dye which upon excitation acts as an intramolecular oxidant.[30] They used constrained DFT-MDs to model the deprotonation reactions, which are expected to take place during the catalytic cycle, as well as the O−O bond formation by a WNA. The formation of the O−O bond together with the subsequent deprotonation of the nucleophile was found to be the energetically most demanding reaction of the catalytic cycle.[30] The work was further extended by introducing a hydroxide as the proton acceptor in the hydration shell. The latter was found to lower the activation barrier for the intermolecular WNA by enhancing the nucleophilicity of the solvent molecule meant to undergo the WNA. Furthermore, by using a hydroxide as the intermolecular base the formation of a high energy hydronium in proximity of the catalyst could be avoided.[31]

Previous work has relied on quite simple order parameters to describe the O−O bond formation, usually the O−O distance. In this work we show how that the use of simple order parameters can lead to a qualitatively wrong picture of the underlying free energy surface. In addition, using a sophisticated amount of solvent molecules in the model system demonstrates the essential role of the solvent environment.

Methods

Computational Settings

All AIMD simulations were performed employing the CP2K program package (revision 18464).[32–34] All atoms were described by the DZVP-MOLOPT-SR-GTH basis sets[35] as well as the corresponding GTH-PBE pseudo potentials.[36] The Perdew-Burke-Ernzerhof (PBE) exchange-correlation functional[37] together with Grimme's D3 dispersion correction,[38] and a cutoff of 800 Ry for the auxiliary plane wave basis set were used to accurately describe the electronic structure. The simulations were performed in the NVT ensemble with a time-step of 0.5 fs, where the temperature was kept constant at 300 K by a Nosé-Hoover chain thermostat.[39,40] The general settings mentioned above closely resemble the protocols

that have been employed previously by us and other groups in the context of enhanced sampling AIMD of transition metal complexes.[41–43] To further underpin the choice of the exchange-correlation functional, the structures of the catalysts optimized in the gas phase were compared with the corresponding single x-ray crystal structure (see Table S1 in the supplementary information).[1] No significant difference among the set of generalized gradient approximation (GGA) exchange-correlation functionals was found. For all simulations a cubic simulation cell with a side length of 14.56 Å, that contains the metal-oxo species of the catalyst ($[Ru^{V}O(Py5OMe)Cl]^{2+}$) and 107 H_2O molecules was employed. Initially, the size of the simulation cell was determined through NPT simulations at 1 bar and 300 K, subsequent equilibration at 300 K in the NVT ensemble resulted in the initial structures for the enhanced sampling calculations.

Bluemoon

A detailed derivation of the *Bluemoon* method can be found in the corresponding original literature by Sprik and Ciccotti.[44–46] Here the basic idea will be introduced. The *Bluemoon* ensemble is, in principle, a thermodynamic integration scheme to calculate the free energy difference (ΔF) between two states. Those are characterized by an order parameter or collective variable (CV) which distinctively separates the two states. For a set of discrete values of the CV (ξ') the average force ($f_{\xi'}$) required to impose the constraint is calculated. Integration of those average forces results in the free energy profile

$$\Delta F = -\int_{\xi_0}^{\xi_1} f_{\xi'} d\xi'. \tag{1}$$

The average force $f_{\xi'}$ is derived from the Lagrange multiplier λ according to

$$f_{\xi'} = \frac{\langle Z^{-1/2}[\lambda - k_{\mathrm{B}}TG]\rangle_{\xi'}}{\langle Z^{-1/2}\rangle_{\xi'}} \tag{2}$$

where k_{B} is the Boltzmann constant, T the temperature, and Z and G are correction factors

associated with the transformation from generalized to Cartesian coordinates. Z is also known as the 'Fixman mass-metric tensor correction' and is defined as

$$Z = \sum_{i=1}^{N} \frac{1}{m_i} \left(\frac{\partial \xi}{\partial \boldsymbol{r_i}} \right)^2 \tag{3}$$

where m_i is the atomic mass of nuclei i, and $\boldsymbol{r_i}$ is the corresponding positional vector. The summation is carried out over all atoms (N) in the system.[47] The G term is defined as

$$G = \frac{1}{Z^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{m_i} \frac{1}{m_j} \frac{\partial \xi}{\partial \boldsymbol{r_i}} \cdot \frac{\partial^2 \xi}{\partial \boldsymbol{r_i} \partial \boldsymbol{r_j}} \cdot \frac{\partial \xi}{\partial \boldsymbol{r_j}}. \tag{4}$$

In the special case where the CV is chosen to be the distance between two atoms i and j ($\xi = r_{ij} = \sqrt{(\boldsymbol{r_i} - \boldsymbol{r_j})^2}$), Z and G simplify to constant values. Therefore, Equation 2 can be rewritten as $f_{\xi'} = \langle \lambda \rangle_{\xi'}$. In all other cases where the order parameter involves multiple distances Z and G have to be derived analytically and evaluated for each step of the simulation.

Because the difference of two distances ($\xi = r_{ij} - r_{jk}$) is used in this work, the analytical expression for Z and G are in this case as follows:

$$Z = \frac{1}{m_i} + \frac{1}{m_i} + \frac{1}{m_k} (1 - \frac{\boldsymbol{r_{ij}}}{r_{ij}} \cdot \frac{\boldsymbol{r_{kj}}}{r_{kj}}) \tag{5}$$

where $\boldsymbol{r_{ij}}$ is the vector difference of the positional vectors of nuclei i and j and

$$G = 0.^{48} \tag{6}$$

Equation 2, thus simplifies to:

$$f_{\xi'} = \frac{\langle Z^{-\frac{1}{2}} \lambda \rangle_{\xi'}}{\langle Z^{-\frac{1}{2}} \rangle_{\xi'}}. \tag{7}$$

Simulations in the *Bluemoon* ensemble were initialized from unconstrained calculations of the R or P state. The reaction path was then defined by a step-wise increase or decrease

of ξ'. It turned out to be crucial that the previous simulation was sufficiently equilibrated before moving to the next step i.e. initialization of all simulations at once was not advisable. For each value of ξ' the simulations were run for about 20 ps (40'000 steps) including an equilibration time of 2.5 ps: Post-processing was performed only on the last 17.5 ps of the trajectory. Note, increasing the equilibration time to 5 ps (10'000 steps) does alter the free energies of the extrema by only $1\,\mathrm{kJ\,mol^{-1}}$ to $2\,\mathrm{kJ\,mol^{-1}}$. This is small with respect to the calculated barrier height.

The standard deviation (σ) of the average force $f_{\xi'}$ is determined by block average methods.[49] By doing so, one obtains an estimate of an upper/lower limit for the free energy difference ΔF using $(f_{\xi'} \pm \sigma_\lambda)$.[50] It is worth mentioning that this is not the only procedure to estimate the standard deviation of the thermodynamic integration of the averages forces, other approaches are known in the literature.[51]

Metadynamics

In the following we point out some key features of MetaD. For a detailed derivation we refer to the original literature by Laio and Parrinello and the corresponding review articles.[52–56]

The MetaD methodology is based on a history dependent bias potential $V(\xi, t)$, which allows the system to leave local minima in order to explore more efficiently the phase space. The bias potential is defined as

$$V(\xi_i, t) = \int_0^t dt' \; \frac{W}{\tau} \; exp\left(- \sum_{i=1}^{d} \frac{(\xi_i - \xi_i(t'))^2}{2\sigma_i^2} \right) \tag{8}$$

where ξ_i is the value of the i-th CV at time t, $\xi_i(t')$ at time t', d are the number of CVs, W the height, τ is the deposition stride, and σ_i is the width of the Gaussian of the i-th CV.

In classical MetaD simulations, an assessment of the convergence is difficult since the exploration of high energy states is enforced through the constant increase of the bias potential until the resulting free-energy is flat, which can occur before the system escapes a

minimum. This problem was circumvented by the introduction of the so called well-tempered metadynamics (WT-MetaD) formalism.[54] In WT-MetaD, the bias deposition rate decreases over the course of the simulation. By doing so, the bias potential will converge to a constant value in the limit of long simulation times. In practice, the reduction of the bias deposition rate is achieved by rescaling the Gaussian height according to

$$W = W_0 \; e^{-\frac{V(\xi,t)}{k_B \Delta T}} \tag{9}$$

where W_0 is the initial height of the Gaussians and ΔT is the temperature difference specifying rescaling rate. Instead of specifying ΔT it is common to define the so-called bias factor γ

$$\gamma = \frac{\Delta T}{T + \Delta T} \tag{10}$$

where T is the temperature of the simulation.

All MetaD simulations were performed using PLUMED (version 2.4.3),[57] together with CP2K (revision 18461).[32] In order to facilitate the sampling, we used six independent walkers all of which contributed to the same metadynamics bias potential.[58] Further, the WT-MetaD formalism together with rigorous error analysis was used to check the convergence of the simulation.[54] Initially, Gaussians with a height of $1\,\mathrm{kJ\,mol^{-1}}$ were added to the bias potential at a pace of 50 steps (i.e. every 25 fs). The bias factor (γ) of the WT-MetaD was set to 25 in order to allow the sampling of barriers with a height of approximately $60\,\mathrm{kJ\,mol^{-1}}$. In addition, depending on the employed CV various restraining potentials were added in order to restrict the sampling space (see description of each individual WT-MetaD simulation in section Metadynamics Simulations). The MetaD simulations have been carried out for **{Ru(Py5OMe)}** employing two different sets of CVs, later referred to as 'set A' and 'set B'. Those simulations will be discussed in detail in section Collective Variable Analysis.

Results and Discussion

In the following sections we will describe the reaction mechanism of the O – O bond formation by a WNA catalyzed by **{Ru(Py5OMe)}** in detail. Employing enhanced sampling techniques such as *Bluemoon* and MetaD with various CVs we show how delicate the choice of a proper method is. In particular, the choice of an appropriate CV is a well known problem in the field of enhanced sampling and various sophisticated protocols such as the committor analysis exist in order to judge the quality of the CV.[53,59,60] Further, there have been attempts to use machine learning and artificial neural networks in order to identify appropriate CVs.[61,62] Here we do not engage in said topic, but rather follow a more chemically guided approach.

Bluemoon Simulations

In our previous study we have investigated a base-assisted mechanism for the O – O bond formation using a static approach.[1,2] However, there can be an alternative pathway where the nucleophile releases a proton to the solvent instead of the base, the later being referred to as the base-independent mechanism. Thus, several solvent molecules are required to stabilize the hydronium or Zundel ion in proximity to the catalyst. For that reason modeling such a mechanism with an approach based on geometry optimizations would have been challenging. Nonetheless, in view of the computational effort required this approach is usually the first choice when it comes to modeling transition states, and there are countless examples in the literature including our own work where several explicit solvent molecules were included in order to model a base-independent reaction mechanism.[8,15–17,63–66] The success of this approach is often dependent on the availability of hydrogen bond acceptor groups within the ligand framework. Those are usually necessary to stabilize the water molecules in proximity to the catalyst which would otherwise explore the whole available conformational space. For the catalyst at hand, besides the dangling pyridine only the chlorido ligand fulfills this cri-

teria. However it is located on the opposite face of the catalyst. This renders a optimization based approach unsuitable. Furthermore, a static approach can lead to serious problems regarding the exact position of the solvent molecules, proper description of the hydrogen bonding network and the localization of the corresponding extrema due to the vast increase in degrees of freedom. These limitations are alleviated in the current model system as the configurational space of both the catalyst and the solvent were explored with AIMD. Notably, under certain experimental conditions the base could have been already protonated prior to the O–O bond formation, which would in principle rule out the base-assisted mechanism.[1] However, this would result in a larger overall charge of the catalysts leading to higher reduction potentials which possibly would prevent the formation of the reactive Ru^{V}=O species. For these reasons, as well as for the sake of consistency with our previous studies, the O–O bond formation is modeled under neutral conditions i.e. none of the solvent molecules are protonated nor is the pyridine.

In line with chemical intuition the most obvious choice for modeling the O–O bond formation by the base-assisted or base-independent mechanisms is a CV based on the distance of the metal-oxo ligand and a specific water molecule $d(O_{oxo}-O_w)$. The two protons of the nucleophile (w) are referred to as H_a and H_b. In this case, both protons are addressed at the same time the abbreviation H_{ab} is used. Free energy profiles obtained by the *Bluemoon* methodology employing a $d(O_{oxo}-O_w)$ CV are shown in Figure 2. Time series of the force acting on the constraint, as well as a linearly interpolated force profile are shown in Figures S2-S4 in the supplementary information.

The larger errors shown in the base-independent pathway (see Figure 2, pale purple line) are partly due to non-uniform sampling of the configurational space of the nucleophile with respect to the base. To further understand this effect we would like to mention that we observed that most simulations describing intermediates between the TS and the R state explore configurations where the minimal distance between the base and the protons of the nucleophile, $(\min(d(N-H_{ab})))$, is larger than 3 Å indicating weak to non existent hydrogen

bonding between the nucleophile and the base (see Figure 3). However, the simulation where $d(O_{oxo}-O_w)$ is constrained to 2.34 Å, explores primarily $\min(d(N-H_{ab}))$ distances around 1.70 Å suggesting hydrogen bonding with the base.

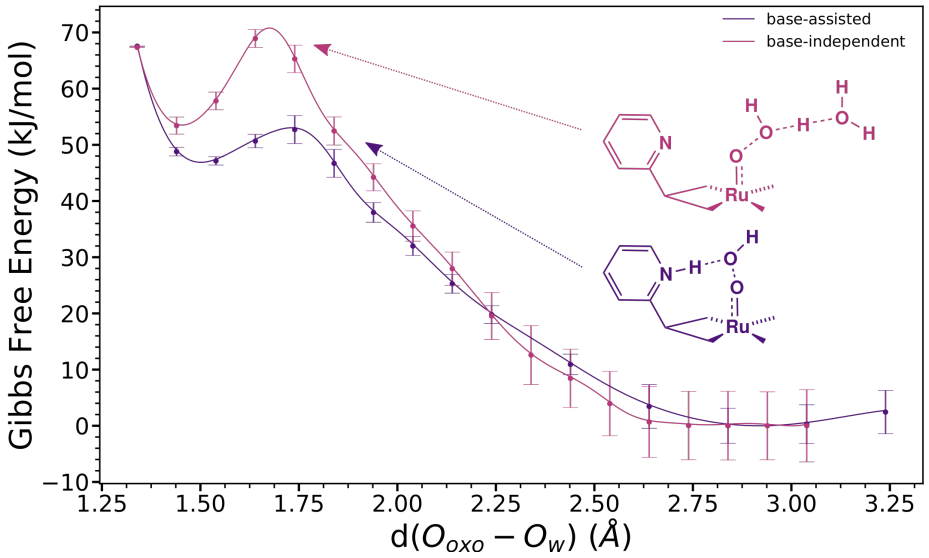


Figure 2: Comparison of the free energy profile for the O−O bond formation by either the base-assisted or the base-independent mechanism. Note that the increase of the error along the reaction path is caused by the integration procedure, where, by convention, the first integration point is set to 0.

Direct comparison of the two reaction mechanisms in terms of the extrema of interest, i.e. R, TS and P, revealed a substantially larger activation barrier of (ΔF_{TS-R}) $69 \pm 8\,\mathrm{kJ\,mol^{-1}}$ in the case of the base-independent mechanism as compared to the $53 \pm 4\,\mathrm{kJ\,mol^{-1}}$ of the base-assisted mechanism (see Table 1). This is likely a consequence of the reduced nucleophilicity of the H_2O molecule, which is destined to undergo the O−O bond formation, and is reflected by a slightly shorter $O_{oxo}-O_w$ distance in the TS (see Figure 2). Identifying the base-assisted mechanism as the energetically most favorable reaction mechanism is in accordance with our previous mechanistic proposal.[1,2]

Since multiple bonds are broken and formed in this reaction utilizing only the $d(O_{oxo}-O_w)$ as a CV might not be a good choice, besides it being the main indication of our desired prod-

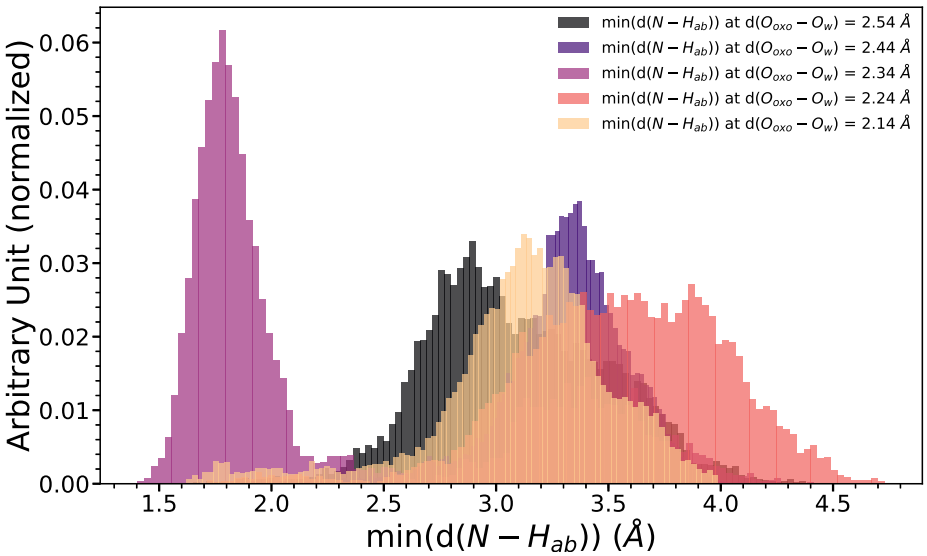


Figure 3: Distribution of the minimum $N-H_{ab}$ distances evaluated for different constrained values of $O_{oxo}-O_w$ distances indicated by the different colors for the base-independent reaction mechanism. The area enclosed by the individual histograms is normalized to unity.

uct. Visualization of the maximal O_w-H_{ab} distance as a function of $d(O_{oxo}-O_w)$ (see Figure 4) revealed that in particular in the region of the transition state ($d(O_{oxo}-O_w)$ = 1.74 Å), there is an abrupt change in the O_w-H_{ab} distance indicating a proton transfer. The lack of intermediate structures describing the proton transfer could potentially be circumvented by increasing the number of constraint simulations between 1.74 and 1.84 Å (see Figure S5 in the supplementary information). However, since the applied CV does not describe the proton transfer explicitly, changing to a different CV is more reasonable.

To this extent, the base-assisted mechanism was modeled using a distance difference CV ($d(O_{oxo}-O_w)$-$d(O_w-H_a)$). The CV space was explored between -0.4 Å and 2.1 Å corresponding to the P and R state respectively, in steps of 0.05 Å in the region of the proton transfer and 0.1 Å elsewhere. A similar CV was used by Sinha *et al.* when simulating the dehydrogenation of alcohols.[67] The obtained free energy profile qualitatively agrees well with the one obtained constraining only the $O_{oxo}-O_w$ distance (see Figure 5 and Table 1). Nonetheless, a stabilization of both the P and the TS by 3 kJ mol^{-1} and 7 kJ mol^{-1} respectively was found,

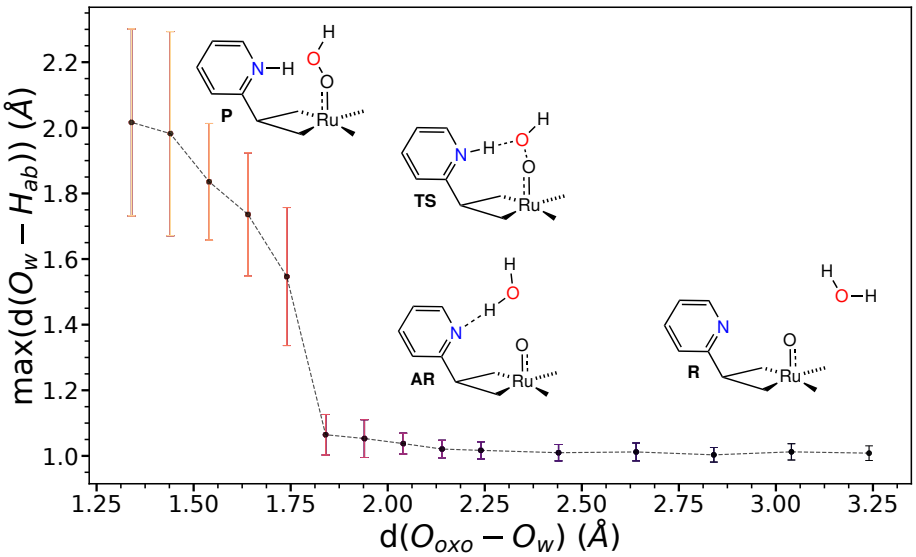


Figure 4: Visualization of max(d(O_w–H_{ab})) and d(O_{oxo}–O_w) distances explored for the base-assisted mechanism. The transition state corresponds to the light red point (5th from the left).

highlighting the importance of including the proton transfer in the CV.

Inspecting the structural features of the TS revealed that in the case of the distance difference CV both the O_{oxo}–O_w and the O_w–H_a distances are slightly elongated (1.78 ± 0.08 Å : 1.58 ± 0.08 Å) as compared to the distance only CV (1.74 Å : 1.5 ± 0.2 Å). This implies that the proton transfer happens prior to the actual O–O bond formation. This is backed up by the better sampled proton transfer, i.e. there is no abrupt proton transfer as in the case of the distance CV only (see Figures 6 and S6 in the supplementary information).

The d(O_{oxo}–O_w)-d(O–H_{ab}) CV is a better descriptor of the base-assisted O–O bond formation than the simpler O_{oxo}–O_w distance only. However, it also comes with a major drawback not discussed up to now. At larger distances from the base the two protons (H_a and H_b) of the nucleophile are chemically indistinguishable. Yet, this is not reflected by the distance difference CV that enforces the release of a specific proton without controlling the spatial orientation of said proton with respect to the base. To do so in a single, chemically intuitive CV appears to be challenging. Therefore the use of a more advanced methodology

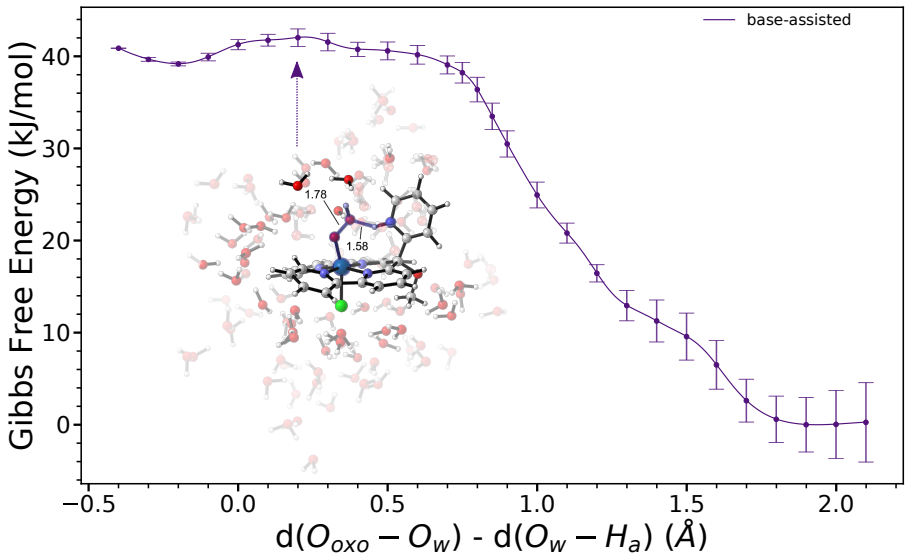


Figure 5: Free energy profile of the base-assisted O−O bond formation catalyzed by **{Ru(Py5OMe)}**, obtained by integrating the average force acting on the constraint (Equation 7) according to Equation 1. The error bars are obtained from block averaging the force profile. Inset: Representative TS structure including the solvation shell. Bond lengths are given in Å.

which allows the use of multiple CVs is advisable.

Table 1: Free energy differences between the R, TS and P highlighting the importance of the base in not only lowering the activation barrier but also in stabilizing the product with respect to the reactant. The magnitude of the standard deviation of the free energy of each individual state depends on its relative location in the reaction path, since by convention the first integration point is set to zero. The errors on the free energy differences (ΔF) are obtained by error propagation and therefore contain contributions from both ends of the reaction path. This makes the direction of the integration from R to P or from P to R less important.

	ΔF_{P-R} (kJ/mol)	ΔF_{TS-R} (kJ/mol)	ΔF_{P-TS} (kJ/mol)
d($O_{oxo}-O_w$) (base-assisted)	47 ± 3	53 ± 4	-6 ± 3
d($O_{oxo}-O_w$) (base-independent)	52 ± 8	69 ± 8	-17 ± 2
d($O_{oxo}-O_w$)-d(O_w-H_a)	39 ± 3	42 ± 3	-3 ± 1

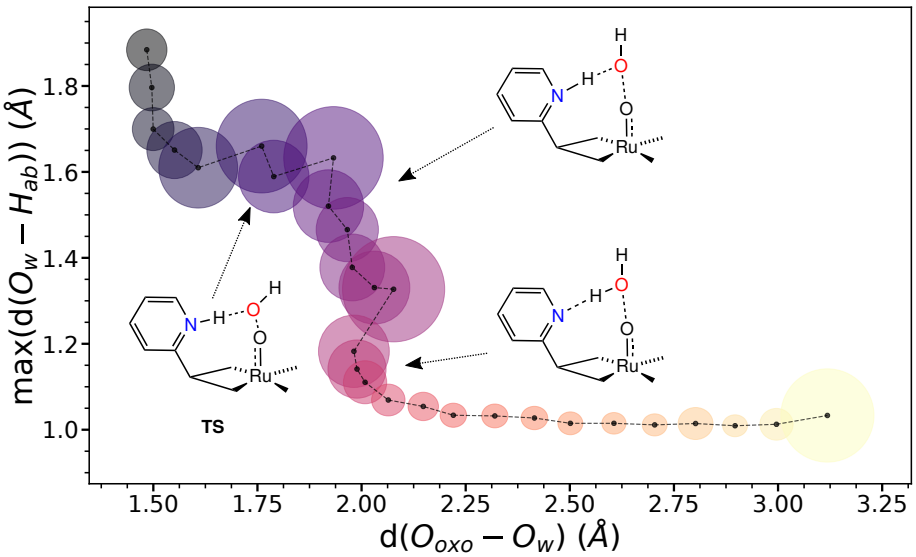


Figure 6: Visualization of the $O_{oxo}-O_w$ and max(O_w-H_{ab}) distances explored during the base-assisted mechanism modeled by employing the distance difference constraint. The width and height of the ellipsis represent the standard deviation of the corresponding CVs. The magnitude of it is approximately inversely proportional to the configurational space explored for each value of the constrained CV i.e. a large standard deviation means that multiple combinations of $O_{oxo}-O_w$ and O_w-Ha distances were explored; as a consequence the number of configurations sampled for each combination alone of d($O_{oxo}-O_w$) and d(O_w-H_{ab}) is small.

Metadynamics Simulations

MetaD simulations offer the option to use multiple different CVs to describe the chemical transformation of interest. Furthermore, additional degrees of freedom might be restrained to prevent the exploration of areas of the phase space that are of less importance. Studies employing either the *Bluemoon* or the MetaD methodology in the context of homogeneous water oxidation, in particular for the O–O bond formation, are scarce. In 2011, Brüssel *et al.* simulated the addition of a CO_2 molecule to a metal coordinated secondary amine, as well as the intramolecular O–O bond formation in the course of a rearrangement of a ruthenium oxo complex, employing both *Bluemoon* and MetaD simulations, where they found a simple distance CV to be insufficient to describe the intermolecular CO_2 addition, a conceptually similar observation as we made in the previous section of this work.[50] Preceding

this work, Vallés-Pardo *et al.* simulated the O–O bond formation by a WNA mechanism, catalyzed by a Ru-based WOC, using the d(O–O) distance as the CV. Since their system did not possess an intramolecular base a solvent molecule acted as the base, accepting a proton from the nucleophile (here referred to as base-independent).[23] However, they did not attempt to observe multiple O–O bond formation events, instead they extracted starting points for simulations in the *Bluemoon* ensemble from the MetaD trajectory, and a strategy has been shown in the previous section, that is hardly applicable to the base-assisted WNA reaction. Piccinin *et al.* used a similar approach where the free energy surface of the O–O bond formation was explored by MetaD. The simulations were already stopped shortly after the first occurrence of the bond formation event and the energetics of the reaction was then determined by single point energy calculations employing a hybrid functional.[24] The subject of their study was a Ru-based polyoxometalate (POM), the structure of which would have allowed an intramolecular O–O bond formation between the Ru=O and an oxygen atom of the POM cage. An initial MetaD simulation rendered the WNA energetically favorable over an intramolecular reaction. In a second MetaD simulation, the WNA was modeled by employing two CVs, one to monitor the coordination number of the oxo-species and thereby the O–O distance and one to keep track of the protonation state of the nucleophile. By doing so they were able to classify the WNA as a concerted reaction consisting of the O–O bond formation and the deprotonation of the nucleophile.[24]

In the following sections we will discuss the application of MetaD simulations to model the O–O bond formation catalyzed by **{Ru(Py5OMe)}**. Thereby, we will take advantage of the insights obtained from the constraint AIMD simulations. Prior to the chemical interpretation of the simulated reaction path, the convergence as well as the choice of an appropriate CV will be assessed.

Choice of the Collective Variables

The most common choices of CVs describing the O–O bond formation have been already discussed in Section Bluemoon Simulations. As has been shown, the base-assisted WNA mechanism could be described appropriately if both the O–O bond formation and the proton transfer are taken into account. This becomes more complex since the indistinguishability of the protons has to be taken into account. In this section, a set of CVs that are able to deal with those difficulties will be introduced.

The indistinguishability of the protons can be resolved by the use of coordination numbers as CVs. Those are defined as

$$CN(r_{ij}) \;=\; \sum_{i}^{M} \frac{\left(1-(\frac{r_{ij}}{r_0})^n\right)}{\left(1-(\frac{r_{ij}}{r_0})^m\right)}, \tag{11}$$

where n and m define the smoothness of the switching function while r_0 specifies the turning point. The summation is carried out for a relevant subset (M) of all the atoms in the system. Usually, n and m are set to 8 and 16 respectively, while r_0 is system dependent. Here $r_0 = 1.5$ Å for set A and $r_0 = 1.3$ Å for set B were used, respectively, as determined by previous trial calculations (see Table 2).

Rather than allowing all the water molecules to act as the nucleophile, we focus the sampling on the main event of the reaction by biasing a selected water molecule. In order to assure that the selected water molecule always remained in close proximity of the reactive site, a quadratic restraining potential was introduced that prevented the O_{oxo}–O_w distance to surpass 2.8 Å (see supplementary information section 'Structural Features of the Extrema' for more details). This distance corresponds to bulk water and was previously determined by simulations (see Section Bluemoon Simulations). By selecting a specific water molecule as the nucleophile, the definition of a second CV was facilitated as the number of indistinguishable protons was reduced to two. We defined the second CV (set A) as the CN of the pyridyl nitrogen with respect to the nucleophile protons. Thus, we use the NH_{ab} coordination pair

rather than the O_wH_{ab} pair. It is worth mentioning that the bias potential applied during the MetaD simulation depends on both CVs. Further a quadratic restraining potential (see Table 2 for details) was used to prevent the formation of a OH^- species in the bulk of the solvent prior to the O–O bond formation, which is unlikely to exist under reaction conditions.[1]

For a second simulation (set B), an even more general description of the proton transfer was used. The second CV in set B takes into account both the protonation state of the nucleophile and the base. This was achieved by defining it as the difference between the coordination number of the base $CN(r_{ij})$ and the nucleophile $CN(r_{kj})$:

$$\Delta CN \;=\; CN(r_{ij}) - CN(r_{kj}) \tag{12}$$

where r_{ij} is the distance of the solvent proton j to the nitrogen atom of the intramolecular base, and r_{kj} the distance of a solvent proton j to the oxygen atom of the nucleophile. By allowing all solvent protons to be involved in the reaction, the reversibility of solvent assisted mechanisms was assured. Monitoring of the protonation state of the nucleophile has already been shown to be a reasonable CV by Piccinin *et al.*[24]

Table 2: Overview of the MetaD simulations settings, CVs and restraining potentials for the sets of simulations A and B.

	Set A		**Set B**	
	description	limits	description	limits
CV1	$d(O_{oxo}–O_w)$	< 2.8 Å	$d(O_{oxo}–O_w)$	< 2.8 Å
CV2	$CN(NH_{ab})$	-	$CN(NH) - CN(O_wH)$	> -2.0
add. restraining pot.	$d(O_w–H_a) - d(N–H_b)$ $- d(O_w–H_b) + d(N–H_a)$	-1.6 Å$< x < 1.6$ Å	-	-

Assessing Convergence

Assessing the convergence of MetaD simulations requires an in-depth analysis of the states sampled and their relative statistics.[53,55] This is in particular true in the context of AIMD where the computational cost is a severe limiting factor.

In general convergence is achieved when the simulation freely transits between the states of interest. This is equivalent to the reoccurring observation of the chemical transformation of interest. The time traces of the CVs in Figure 7 show multiple O−O bond formations and breakages as well as the protonation and deprotonation events of the base within the same walker and across multiple walkers. When the free energy difference between the reactant and product is not of interest, it is common to stop MetaD simulations after the first occurrence of the reaction of interest.[68] In this work, we intentionally went beyond this simplification in order to get a full picture of both the reactant and product states and their free energy difference.

Another indication of convergence can be obtained by projecting the free energy surfaces (FESs) onto the degrees of freedom that were biased i.e. in the case of set B the $O_{oxo}-O_w$ distance and the difference of the coordination numbers (CN(NH)-CN(O_wH)). An estimate of the standard deviation of the free energy profile is shown in Figure 8. It was obtained by calculating block averages (5 blocks) over the concatenated trajectory of the six individual walkers (see also Figure S7 in the supplementary information). In general the standard deviation was found to be within a few kJ/mol indicating uniform sampling of the whole CV space. Note the variable magnitude of error over the CV space indicates different degrees of convergence. This is especially true for the upper and lower limit of the explored CV space where the sampling is poor.

The convergence was further validated by calculating an estimated error on the MEPs connecting the R and P state (see Section Characteristics of the FES). Applying block averaging methods an optimal block number of 5 was determined for the error on the MEP (see Figure S8 in the supplementary information). Note the standard deviations given here only account for the statistical error of the sampling, not the systematical errors of DFT caused by e.g. the choice of the exchange-correlation functional and basis set.

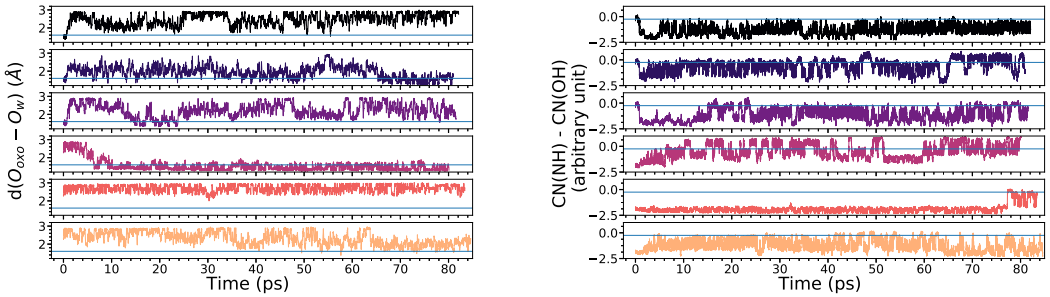


Figure 7: Left: Time-trace of the O–O distance for different walkers of **{Ru(Py5OMe)}** (set B). The blue horizontal line indicates the formation of the O–O bond. Right: Time-trace of CN(NH) - CN(OH) of the same simulation. Formation of the hydroperoxo species is indicated by the horizontal blue line. Note the horizontal lines merely serve as a visual guide line rather than a strict assignment of states.

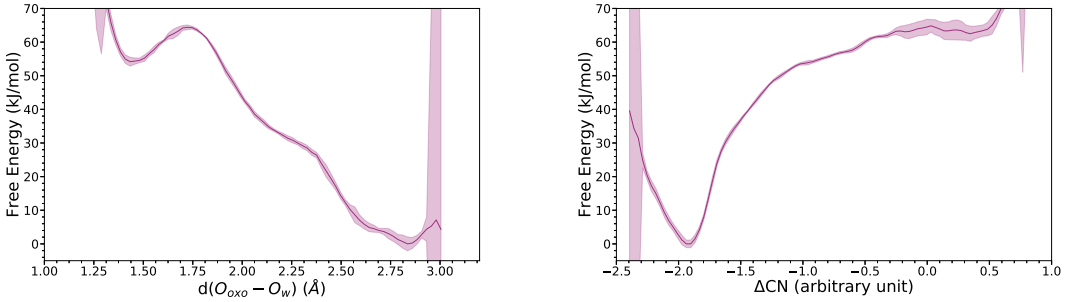


Figure 8: Left: Free energy profile including the error range of the O_{oxo}–$O_w$distance for **{Ru(Py5OMe)}** (set B). Right: Free energy profile including the error range of the CN(NH) – CN(O_wH) CV of the same simulation. Note the free energy profiles shown here were obtained by reweighting procedures accounting for the additional bias imposed by the restraining potentials.

Collective Variable Analysis

The CVs biased in both set A and set B successfully describe the O–O bond formation by a base-assisted mechanism, which can be seen by the exploration of the transition state region (set A: CV1 ≈ 1.8 Å, CV2 ≈ 0.9 and set B: CV1 ≈ 1.7 Å, CV2 ≈ -0.2) (see Figure 9). There were significant differences with respect to both reaction pathways and states explored between the two different sets. This can be seen in Figure 9 where the simulations biasing the CVs from set A were reweighted according to the CVs from set B and vice versa.

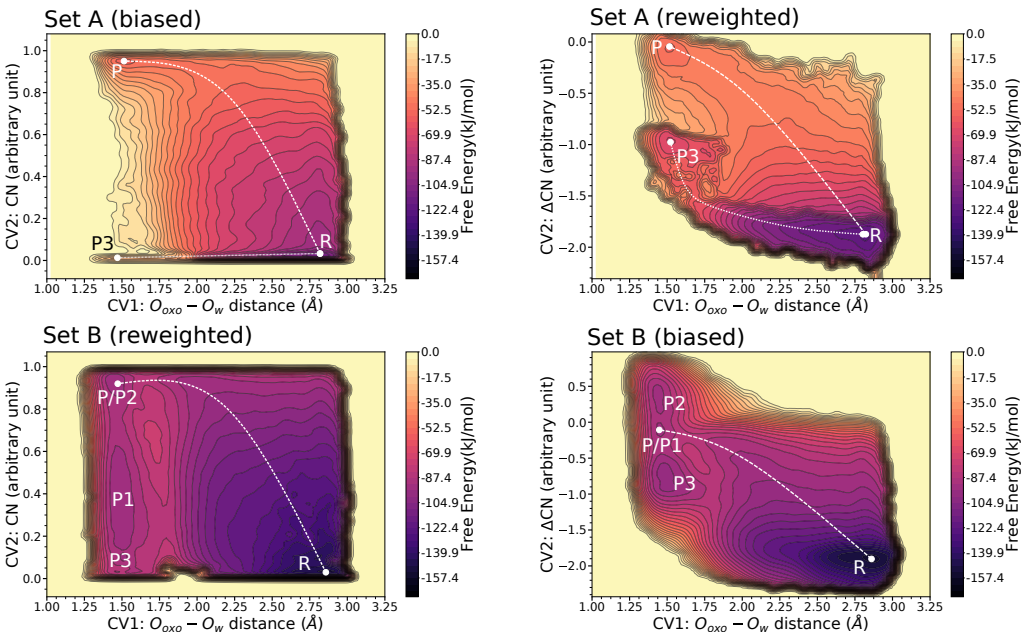


Figure 9: Top-left: FES of **{Ru(Py5OMe)}** reconstructed using the settings of set A. Top-right: Reweighted FES (set A) according to CVs of set B. Bottom-left: Reweighted FES (set B) of **{Ru(Py5OMe)}** according to CVs of set A. Bottom-right: FES of **{Ru(Py5OMe)}** reconstructed using the settings of set B. The dashed line is a visual aid to indicate the base-assisted pathway, analogously the base-independent is indicated by the dotted line.

The first notable difference amongst the sets of CVs is that the base-independent reaction path, directly connecting the R and P3 state (see Figure 10), only occurs when the CVs from set A are biased by MetaD, which is a key difference between the two simulations (see Figure 9, top-left and top-right panel). Thereby, a proton is released to the nearest solvent molecule and from there transfered by the Grotthuss mechanism to different solvent molecules. In the case of set A, the protons to be transfered to the base are explicitly defined. Due to this, if one of those protons hops to the solution there is no direct bias bringing it back to the reaction region, consequently the reaction halts and the simulation is then stuck into sampling product like states, which compromises the overall statistics. The possibility of a base-independent reaction in the case of set B will be discussed in detail in section Characteristics of the FES.

The second striking difference is the co-existence of three energetically similar local min-

ima in which the O−O bond has been formed in the case of the simulation employing the CVs of set B (see Figure 9, bottom-right panel). Those are states with either a formed hydroperoxo species and the protonated base (P), a peroxo species and a protonated base where a proton was released to the solvent (P2) or a hydroperoxo species with a deprotonated base where the proton of the nucleophile was released to the solvent (P3) (see Figure 10). The P3 state corresponds to the product of the base-independent pathway, but can also be formed via proton transfer from the hydroperoxo ligand of P to the solvent or by the deprotonation of its base. A possible intermediate structure where a proton is shared between the base and the peroxo ligand (P1) is indistinguishable from the P state within the CVs from set B. The absence of those species in the case of set A is likely a consequence of the applied restraining potential on the $d(O_w - H_{ab})$ and $d(N - H_{ab})$ distances. This potential was introduced based on previous trial runs in order to keep the 'active' protons (H_a and H_b) in between the base and the nucleophile. The spontaneous deprotonation of the hydroperoxo ligand has recently also been observed for Ru-WOC linked to a dye molecule by Shao *et al.*[31]

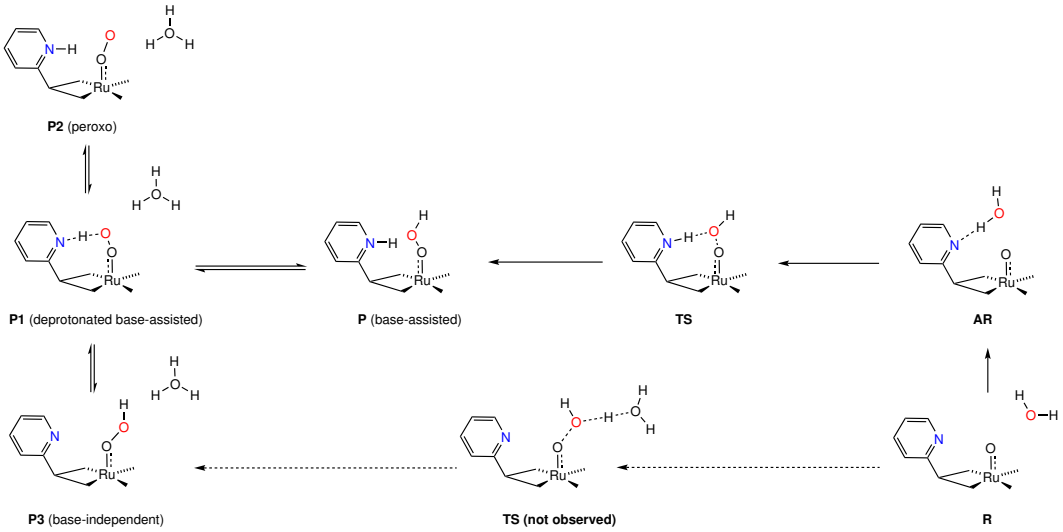


Figure 10: Schematic overview of the reaction network connecting the observed states. Note that P and P1 co-inhabit the same free energy basin and thus are virtually indistinguishable within the chosen CVs.

Characteristics of the FES

MEPs were identified on the smoothed FES using the minimum energy pathway analysis for energy landscapes (MEPSA) employing a variation of the Dijkstra's algorithm.[69,70] MEPs obtained by this procedure by no means have an exclusive character i.e. they are the lowest energy pathways connecting two predefined local minima, but there might be several alternative pathways that are energetically similar, i.e. that have the same or a slightly higher energy cost associated. By evaluating the final MEP on the FESs obtained for the individual blocks used for the block analysis an estimate of its error can be given (see Figure 13).

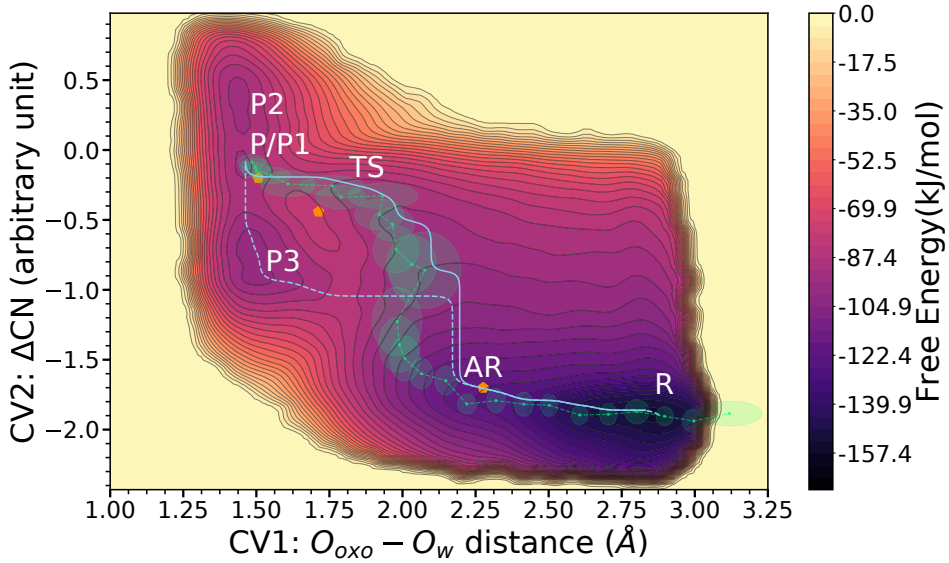


Figure 11: Visualization of the MEP for the O−O formation on the FES of **{Ru(Py5OMe)}** (set B) following the base-assisted mechanism (solid line) and the base-independent mechanism (dashed line). The base-assisted reaction path sampled within the *Bluemoon* ensemble projected onto the same FES is represented by the pale green dashed line and the ellipses. The height and width of the ellipses correspond to the standard deviation of the CVs calculated from the trajectories of the constrained simulations. The orange pentagons correspond to the AR, TS and P state obtained by the static DFT simulations reported by Gil-Sepulcre *et al.*[1]

The inherent complexity of this process translates into an equally intricate FES making the unambiguous identification of the MEP difficult (see Figure 11). In order to differentiate

the two pathways, the FES was projected into a third dimension (see Figure 12). As shown in Figure 9, following the formation of the N−H bond can directly monitor the contribution of the base to the overall reaction, thus identifying a clear separation amongst the two possible reaction pathways. Conceptually similar representations have been used in cases where more than two CVs were required to model the reaction.[71] The FES clearly shows how the local minima representing the R and the P state are connected trough a low energy transition region, highlighting the fact that the transition region was exhaustively explored during the simulation. On the other hand, there is no low energy region connecting the R and the P3 state as it would be required in the case of the base-independent pathway. The three dimensional representation of the FES in Figure 12 further implies that, as soon as P is formed, proton transfer from the base to the hydroperoxo ligand and from the hydroperoxo to the solvent ligand becomes energetically feasible. Therefore, the formation of the O−O bond without the involvement of the base, as suggested by the dashed-MEP (see Figure 11) could be excluded leaving the base-assisted MEP as the only viable option. This is in agreement with the *Bluemoon* simulations discussed in section Bluemoon Simulations. A set of exemplary structures along the preferred base-assisted MEP is shown in Figure 13. During the first phase of the reaction, the nucleophile is in close proximity to the catalyst, then it gets in a proper orientation for the O−O bond formation by engaging in hydrogen-bonding with the intramolecular base. Upon a successful proton-transfer, the distance between the nucleophile and oxo-ligand is reduced, leading the proposed TS. Continuing from there the O−O bond is then finally formed. During the whole reaction several solvent molecules engage in hydrogen-bonding with the nucleophile. This is particularly pronounced for intermediate states where the proton is shared between the nucleophile and the base.

The base-assisted MEP obtained from the MetaD biasing the CVs from set B was further verified by comparing it with the MEP obtained on the reweighted FES (biased CVs from set B, but reweighted according to the set A CVs). The free energy differences between the extrema (R, TS and P) of those simulations are given in Table 3. For the MetaD simulation

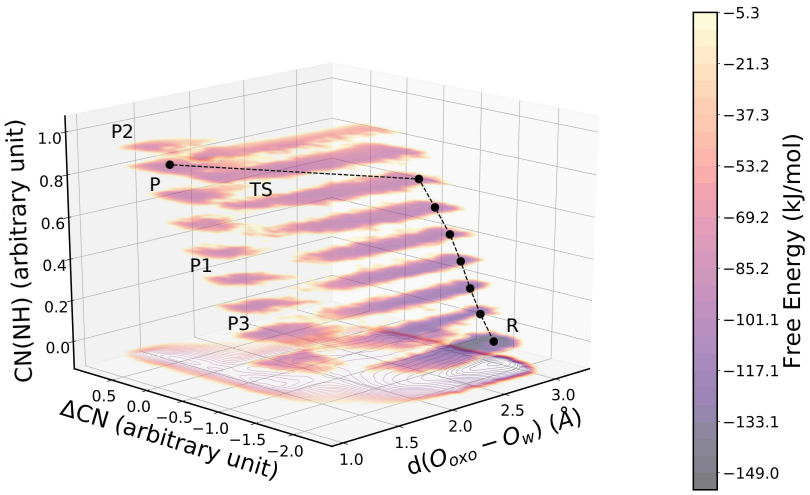


Figure 12: FES of **{Ru(Py5OMe)}** reconstructed using the MetaD simulations with the set B of CVs (x- and y-axis). Additionally, a third dimension, the coordination number of the base (CN(NH)) (z-axis), is used to visualize the preferred reaction path. Note that the path shown in the graph only serves as a visual aid and does only qualitatively correspond to the MEP shown previously.

biasing set B, as well for set B reweighted according to set A the relative energies are qualitatively the same. Naturally, reweighting results in larger errors, in particular for the reactant state which due to the generalized definition in set B (i.e. the coordination number of all protons instead of two specific ones) is slightly different among the two sets of CVs. The larger errors associated with set A highlight the fact that the simulation did not reach the same sampling efficiency as one of set B, especially due to the reversibility of the hydronium formation as discussed in Section Collective Variable Analysis. Note, since the simulation employing the CVs of set A got stuck, its length is only about half of the simulation using the CVs of set B. Nonetheless, qualitatively stability of the reactant, transition state and product are recovered.

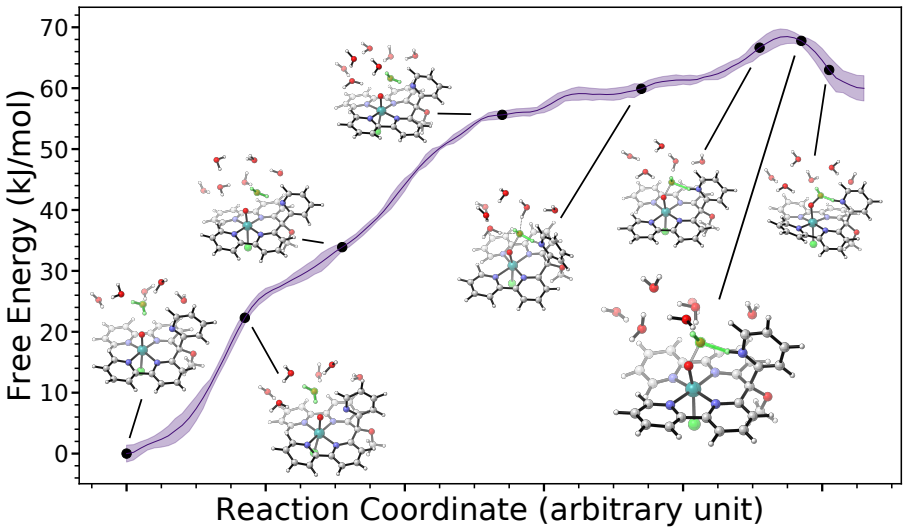


Figure 13: MEP of the base-assisted O – O bond formation of **{Ru(Py5OMe)}** obtained from the MetaD simulations with the set B of CVs. An estimation of the error of the MEP is shown together with exemplary structures along the reaction coordinate. Those were obtained by the clustering algorithm implemented in METAGUI 3.0.[72]

Table 3: Free energy differences obtained from the MEP of the FES. The reweighted pairs highlight the different sampling efficiency amongst the CV sets. All values given are in kJ/mol.

	{Ru(Py5OMe)}		
	ΔF_{P-R}	ΔF_{TS-R}	ΔF_{P-TS}
MEP (set A)	59 ± 48	64 ± 42	-6 ± 24
MEP (set A, reweighted B)	70 ± 21	77 ± 17	-7 ± 26
MEP (set B)	60 ± 3	69 ± 2	-8 ± 3
MEP (set B, reweighted A)	55 ± 9	65 ± 9	-10 ± 3

Confronting Different Computational Methodologies

In our previous study, we applied static DFT simulations to model the O – O bond formation, whereby our model system contained two explicit water molecules, one of which acted as the nucleophile. In the TS structure, the additional water molecules are engaged in hydrogen bonding with the nucleophile locking it in place for the O – O bond formation. Lack of additional solvent molecules with whom hydrogen bonding would have been possible resulted in the AR structure that closely resembled the TS. As a consequence, the energetics of the

AR, TS and P were referenced against the corresponding Ru^{V}=O species plus two H_2O, obtained by separate optimizations, as they better represent the energetics of the R state.[1,2] Projecting the AR, TS and P states reported in the earlier study onto the FES obtained by the MetaD simulations (see Figure 11) reveals that neither the AR nor the P are local minima on the MetaD FES.[2] This comes as no surprise since the underlying electronic structure method as well as the description of the surrounding solvent is different from the current MetaD simulations. A detailed comparison of key structural features of the extrema describing the O−O bond formation can be found in Section Structural Features of the Extrema in the supplementary information. In the case of the static DFT simulations geometries were optimized at the BP86-D3/def2-TZVP level of theory within the TURBOMOLE program package.[73–76] The reported free energies were then obtained using single point calculations with B3LYP-D3/def2-TZVP and the rigid rotator harmonic oscillator approach,[77,78] whereby the interactions with the solvent were approximated by COSMO. Those values were further corrected for ambient temperature and pressure (see Ref.[1] for more details).[1,2] Nevertheless, the structures are similar. In particular, the structure of the TS and P are located in close proximity to their correspondent counterparts on the FES obtained by MetaD simulations. Further, the AR lies approximately on the MEP connecting the R and P on the FES and is clearly different from bulk water supporting our choice of referencing the free energies against the Ru^{V}=O species.

When evaluating the free energies of the AR, TS, and P states obtained by static calculations on the FES a surprising agreement between the free energies of the TS was found, despite the vast differences in the applied methodologies (see Table 4). On the other hand, substantial differences in the stability of the P state were found, which are likely attributed to the fact that the static DFT structures are locked in a single configuration by hydrogen bonding of the hydroperoxo ligand and the protonated intramolecular base ($d(O_{HOO}-H)$= 1.69 Å (static) vs. 2.0 ± 0.2 Å (MetaD)) and the one additional solvent molecule, resulting in the stabilization of the P state. In the case of MetaD simulations,

the existence of the product like states where the solvent has been protonated suggests that under ambient conditions the tightly locked conformation found in static simulations is energetically unfavorable.

Here we want to point out, that the stability of those intermediates containing a hydronium in solution is likely to be underestimated due to the fact that there might be spurious interactions between the two positively charged species in the simulation cell. This could be alleviated by using a significantly large simulation cell or by including a negatively charged proton acceptor such as a hydroxide in the solvation shell. Both options are beyond the scope of the current work where we focused on the O−O bond formation.

When describing the reactant using DFT geometry optimization, weak interactions between the base and the metal-oxo might additionally stabilize the reactant state. This interaction is best described by the d(N−O_{oxo} = 2.4 Å) distance and the angle of the pyridyl with respect to the plan spanned by the Ru metal center and the ligand ∠($C_{py}C_{OMe}$Ru) = 100.1 ° (see Figure S1 in the supplementary informations). Within the enhanced sampling methods used in this work, this distance increases to d(N−O_{oxo}) = 2.7 ± 0.3 Å and ∠($C_{py}C_{OMe}$Ru) to 104 ± 5 ° implying an even weaker interaction between the Ru and the pyridyl but no complete absence of interactions. A strong pyridyl-oxo interaction would be an indication that the formation of a pyridyl-N-oxide by an oxygen atom transfer could be energetically favorable. The absence of the latter renders such a reaction pathway less likely. However, the verification of this interpretation might be the topic of further studies.

Table 4: Free energies of the AR, TS and P structures reported by Gil-Sepulcre *et al.* and their corresponding relative value when projected onto the FES (set B). The free energies of the MetaD simulations are given relative to the local minima corresponding to bulk water. All free energies are given in kJ/mol.[1]

	{Ru(Py5OMe)}			
	R	AR	TS	P
static DFT[2]	0	29	65	26
static DFT projected on FES	0 ± 2	31 ± 2	71 ± 2	63 ± 3
MetaD (set B)	0 ± 2	-	69 ± 2	62 ± 3

Analogously, the reaction pathways obtained by the *Bluemoon* simulations using the difference of distances constraint and the MetaD simulations were compared by projecting the respective MEPs onto the FES (set B) (see Figure 11). The reaction path explored by the *Bluemoon* simulation follows the same features of the FES as the MEP, thereby visiting similar extrema i.e. the R, the TS and the P. Consequentially, the free energies of the projected extrema were almost identical to the extrema visited by the MEP (see Table 5). A direct comparison of the relative stabilities of extrema suggested that within the *Bluemoon* ensemble the reactant state is destabilized by almost 20 kJ compared to the MetaD simulation. The difference is likely a consequence of the rather simple CVs used within the *Bluemoon* ensemble. The mean force therefore lacks significant contributions from other degrees of freedom relevant for the O−O bond formation.[50] Those other degrees of freedom appear to be in particular important for the description of the proton transfer. A further limitation of the CV arising from the explicit specification of the 'active' proton was already discussed in section Bluemoon Simulations.

Table 5: Comparison the energetics of the R, TS and P states obtained with *Bluemoon*, MetaD and the projection of the *Bluemoon* path onto the MetaD FES. All free energies are in kJ/mol.

	{Ru(Py5OMe)}		
	R	TS	P
Bluemoon	0 ± 2	42 ± 3	39 ± 3
Bluemoon path projected on FES	0 ± 2	69 ± 3	61 ± 3
MetaD (set B)	0 ± 2	69 ± 2	62 ± 3

Conclusion

Applying forefront *ab initio* molecular dynamics in combination with enhanced sampling methods, we investigated in detail the O−O bond formation catalyzed by the **{Ru(Py5OMe)}** WOC. Using a large number of explicit solvent molecules in a periodic framework treated at ambient conditions and at the DFT level allows for a sophisticated

description of the solvent and finite temperature effects and is thus a comprehensible approach on how to improve the picture obtained by static DFT calculations. We showed that a base-assisted mechanism is the energetically favorable pathway for the O – O bond formation by a WNA contrary to a base-independent mechanism. This was achieved in two ways, first by forcing the exploration of different pathways within the *Bluemoon* ensemble, and second by sampling pathways in an unrestrained manner utilizing the MetaD simulations. In particular, modeling of the base-independent mechanism would have been challenging within a static solvation approach as used in previous work, because the stabilization of a hydronium in proximity to the catalyst would require several additional solvent molecules. The inclusion of which would require the use of sampling techniques otherwise the results would become ambiguous.

Focusing on the base-assisted mechanism, we have shown how the interplay between the base and the nucleophile results in the formation of a hydrogen-bonded reactive intermediate prior to the O – O bond formation. A proton transferred to the base subsequently leads to the formation of the O – O bond. The refinement of the CVs choice used to monitor the O – O bond formation demonstrated that, being in line with chemical intuition, simple reaction coordinates can lead to the exploration of unexpected areas of the FES, but may miss relevant features of the process of interest.

A direct comparison between energetics of the reaction path simulated by either the static, *Bluemoon*, or MetaD model is usually not expedient due to the differences in the methods. However, there was still a remarkable agreement among the structures of the local extrema (R, TS, and P). The energetics obtained in that manner was in good agreement among the three different methodologies. To the best of our knowledge, this study is the first direct comparison of those different simulation protocols in the context of a O – O bond formation by a WNA and is therefore important for the emerging field of water splitting for sustainable energy storage and conversion.

The biggest qualitative differences between theenergetics obtained by the static DFT

simulations and the ones obtained by enhanced sampling AIMDs were related to the product state, which turned out to be less stable than the suggested by geometry optimizations. On the other hand, the instability led to a variety of product states that were isoenergetic, among them there were deprotonated species which represent the expected intermediates further down the catalytic cycle, such as the peroxo-species (P1) suggesting a high reactivity of the P. The detection of such isoenergetic states are a clear indication of how necessary a robust sampling is, since it can unveil new system capabilities which can be paramount in other applications.

Overall, we have shown that modeling of the O−O bond formation by a WNA within a static DFT framework and asserted limitations due to the description of the solvent dynamics can lead to an oversimplified picture of the reaction of interest and the reaction network as well as potentially wrong estimates of the reaction barrier. Enhanced sampling methods such as *Bluemoon* and MetaD offer a suitable solution to proper sample solvent dynamics and finite temperature effects while having many explicit solvent molecules. However, they come at a significant cost. Nevertheless, this then allows unprecedented insight in the process studied and the relevant parts of the phase space.

Acknowledgment

The work has been supported by the University of Zurich, the university research priority program 'Solar Light to Chemical Energy Conversion' (LightChEC) and the Swiss National Science Foundation (grant no. PP00P2_170667). We thank the Swiss National Supercomputing Center for computing resources (project ID: s745 and s788). MS is further grateful for the discussions with Momir Mališ.

References

(1) Gil-Sepulcre, M.; Böhler, M.; Schilling, M.; Bozoglian, F.; Bachmann, C.; Scherrer, D.; Fox, T.; Spingler, B.; Gimbert-Suriñach, C.; Alberto, R.; Bofill, R.; Sala, X.; Luber, S.; Richmond, C. J.; Llobet, A. Ruthenium Water Oxidation Catalysts based on Pentapyridyl Ligands. *ChemSusChem* **2017**, *10*, 4517–4525.

(2) Schilling, M.; Böhler, M.; Luber, S. Towards the rational design of the Py5-ligand framework for ruthenium-based water oxidation catalysts. *Dalton Trans.* **2018**, *47*, 10480–10490.

(3) Lai, W.; Cao, R.; Dong, G.; Shaik, S.; Yao, J.; Chen, H. Why Is Cobalt the Best Transition Metal in Transition-Metal Hangman Corroles for O−O Bond Formation during Water Oxidation? *J. Phys. Chem. Lett.* **2012**, *3*, 2315–2319.

(4) Li, X.; Siegbahn, P. E. M. Water oxidation mechanism for synthetic Co-oxides with small nuclearity. *J. Am. Chem. Soc.* **2013**, *135*, 13804–13813.

(5) Tong, L.; Duan, L.; Xu, Y.; Privalov, T.; Sun, L. Structural Modifications of Mononuclear Ruthenium Complexes: A Combined Experimental and Theoretical Study on the Kinetics of Ruthenium-Catalyzed Water Oxidation. *Angw. Chem. Int. Ed.* **2011**, *50*, 445–449.

(6) Schilling, M.; Luber, S. Computational Modeling of Cobalt-Based Water Oxidation: Current Status and Future Challenges. *Front Chem* **2018**, *6*, 100.

(7) Shaffer, D. W.; Xie, Y.; Concepcion, J. J. O−O bond formation in ruthenium-catalyzed water oxidation: single-site nucleophilic attack vs. O−O radical coupling. *Chem. Soc. Rev.* **2017**, *46*, 6170–6193.

(8) Liao, R.-Z.; Siegbahn, P. E. M. Quantum Chemical Modeling of Homogeneous Water Oxidation Catalysis. *ChemSusChem* **2017**, *10*, 4236–4263.

(9) Vogiatzis, K. D.; Polynski, M. V.; Kirkland, J. K.; Townsend, J.; Hashemi, A.; Liu, C.; Pidko, E. A. Computational Approach to Molecular Catalysis by 3d Transition Metals: Challenges and Opportunities. *Chem. Rev.* **2019**, *119*, 2453–2523.

(10) Luber, S. Advancing Computational Approaches for Study and Design in Catalysis. *CHIMIA* **2018**, *72*, 508–513.

(11) Harvey, J. N.; Himo, F.; Maseras, F.; Perrin, L. Scope and Challenge of Computational Methods for Studying Mechanism and Reactivity in Homogeneous Catalysis. *ACS Catal.* **2019**, *9*, 6803–6813.

(12) Schilling, M.; Luber, S. In *Water Oxidation Catalysts*; van Eldik, R., Hubbard, C. D., Eds.; Adv. Inorg. Chem.; Academic Press, 2019; Vol. 74; pp 61–114.

(13) Schäfer, A.; Klamt, A.; Sattel, D.; Lohrenz, J. C. W.; Eckert, F. COSMO Implementation in TURBOMOLE: Extension of an efficient quantum chemical code towards liquid systems. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2187–2193.

(14) Klamt, A. The COSMO and COSMO-RS solvation models. *WIREs Comput Mol Sci.* **2011**, *1*, 699–709.

(15) Hodel, F. H.; Luber, S. What Influences the Water Oxidation Activity of a Bioinspired Molecular $Co^{II}{}_4O_4$ Cubane? An In-Depth Exploration of Catalytic Pathways. *ACS Catal.* **2016**, *6*, 1505–1517.

(16) Hodel, F. H.; Luber, S. Redox-Inert Cations Enhancing Water Oxidation Activity: The Crucial Role of Flexibility. *ACS Catal.* **2016**, *6*, 6750–6761.

(17) Scherrer, D.; Schilling, M.; Luber, S.; Fox, T.; Spingler, B.; Alberto, R.; Richmond, C. J. Ruthenium water oxidation catalysts containing the non-planar tetradentate ligand, biisoquinoline dicarboxylic acid (biqaH2). *Dalton Trans.* **2016**, *45*, 19361–19367.

(18) Hodel, F. H.; Deglmann, P.; Luber, S. Exploring Solvation Effects in Ligand-Exchange Reactions via Static and Dynamic Methods. *J. Chem. Theory Comput.* **2017**, *13*, 3348–3358.

(19) Blumberger, J.; Ensing, B.; Klein, M. L. Formamide Hydrolysis in Alkaline Aqueous Solution: Insight from Ab Initio Metadynamics Calculations. *Angw. Chem. Int. Ed.* **2006**, *45*, 2893–2897.

(20) Sala, O.; Lüthi, H. P.; Togni, A.; Iannuzzi, M.; Hutter, J. Dividing a complex reaction involving a hypervalent iodine reagent into three limiting mechanisms by ab initio molecular dynamics. *J. Comput. Chem.* **2015**, *36*, 785–794.

(21) Sala, O.; Santschi, N.; Jungen, S.; Lüthi, H. P.; Iannuzzi, M.; Hauser, N.; Togni, A. S-Trifluoromethylation of Thiols by Hypervalent Iodine Reagents: A Joint Experimental and Computational Study. *Chem. Eur. J.* **2016**, *22*, 1704–1713.

(22) Liu, L.; Lukose, B.; Ensing, B. A Free Energy Landscape of CO_2 Capture by Frustrated Lewis Pairs. *ACS Catal.* **2018**, *8*, 3376–3381.

(23) Vallés-Pardo, J. L.; Guijt, M. C.; Iannuzzi, M.; Joya, K. S.; de Groot, H. J. M.; Buda, F. Ab Initio Molecular Dynamics Study of Water Oxidation Reaction Pathways in Mono-Ru Catalysts. *Chem. Phys. Chem.* **2012**, *13*, 140–146.

(24) Piccinin, S.; Sartorel, A.; Aquilanti, G.; Goldoni, A.; Bonchio, M.; Fabris, S. Water oxidation surface mechanisms replicated by a totally inorganic tetraruthenium–oxo molecular complex. *Proc Natl Acad Sci USA* **2013**, *110*, 4917–4922.

(25) Govindarajan, N.; Tiwari, A.; Ensing, B.; Meijer, E. J. Impact of the Ligand Flexibility and Solvent on the O-O Bond Formation Step in a Highly Active Ruthenium Water Oxidation Catalyst. *Inorg. Chem.* **2018**, *57*, 13063–13066.

(26) Warshel, A.; Weiss, R. M. An empirical valence bond approach for comparing reactions in solutions and in enzymes. *J. Am. Chem. Soc.* **1980**, *102*, 6218–6226.

(27) Zhan, S.; De Gracia Triviño, J. A.; Ahlquist, M. S. G. The Carboxylate Ligand as an Oxide Relay in Catalytic Water Oxidation. *J. Am. Chem. Soc.* **2019**, *141*, 10247–10252.

(28) Pushkar, Y.; Pineda-Galvan, Y.; Ravari, A. K.; Otroshchenko, T.; Hartzler, D. A. Mechanism for O−O Bond Formation via Radical Coupling of Metal and Ligand Based Radicals: A New Pathway. *J. Am. Chem. Soc.* **2018**, *140*, 13538–13541.

(29) Govindarajan, N.; Meijer, E. J. Modeling the Catalyst Activation Step in a Metal–Ligand Radical Mechanism Based Water Oxidation System. *Inorganics* **2019**, *7*.

(30) Shao, Y.; de Ruiter, J. M.; de Groot, H. J. M.; Buda, F. Photocatalytic Water Splitting Cycle in a Dye-Catalyst Supramolecular Complex: Ab Initio Molecular Dynamics Simulations. *J. Phys. Chem. C* **2019**, *123*, 21403–21414.

(31) Shao, Y.; de Groot, H. J.; Buda, F. Proton Acceptor near the Active Site Lowers Dramatically the O–O Bond Formation Energy Barrier in Photocatalytic Water Splitting. *J. Phys. Chem. Lett.* **2019**, *10*, 7690–7697.

(32) CP2K Developers Group, CP2K Program Package. `https://www.cp2k.org/`.

(33) VandeVondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J. Quickstep: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Comput. Phys. Commun.* **2005**, *167*, 103–128.

(34) Hutter, J.; Iannuzzi, M.; Schiffmann, F.; VandeVondele, J. Cp2k: Atomistic simulations of condensed matter systems. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2014**, *4*, 15–25.

(35) VandeVondele, J.; Hutter, J. Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases. *J. Chem. Phys.* **2007**, *127*, 114105.

(36) Goedecker, S.; Teter, M.; Hutter, J. Separable dual-space Gaussian pseudopotentials. *Phys. Rev. B* **1996**, *54*, 1703–1710.

(37) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(38) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.

(39) Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **1984**, *81*, 511–519.

(40) Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **1984**, *52*, 255–268.

(41) Koitz, R.; Iannuzzi, M.; Hutter, J. Building Blocks for Two-Dimensional Metal–Organic Frameworks Confined at the Air–Water Interface: An Ab Initio Molecular Dynamics Study. *J. Phys. Chem. C* **2015**, *119*, 4023–4030.

(42) Koitz, R.; Hutter, J.; Iannuzzi, M. Formation and properties of a terpyridine-based 2D MOF on the surface of water. *2d Mater.* **2016**, *3*, 025026.

(43) Song, F.; Moré, R.; Schilling, M.; Smolentsev, G.; Azzaroli, N.; Fox, T.; Luber, S.; Patzke, G. R. $\{Co_4O_4\}$ and $\{Co_xNi_{4-x}O_4\}$ Cubane Water Oxidation Catalysts as Surface Cut-Outs of Cobalt Oxides. *J. Am. Chem. Soc.* **2017**, *139*, 14198–14208.

(44) Sprik, M.; Ciccotti, G. Free energy from constrained molecular dynamics. *J. Chem. Phys.* **1998**, *109*, 7737–7744.

(45) Ciccotti, G.; Ferrario, M. Blue Moon Approach to Rare Events. *Mol. Sim.* **2004**, *30*, 787–793.

(46) Ciccotti, G.; Kapral, R.; Vanden-Eijnden, E. Blue Moon Sampling, Vectorial Reaction Coordinates, and Unbiased Constrained Dynamics. *ChemPlusChem* **2005**, *6*, 1809–1814.

(47) den Otter, W. K. Revisiting the Exact Relation between Potential of Mean Force and Free-Energy Profile. *J. Chem. Theory Comput.* **2013**, *9*, 3861–3865.

(48) Komeiji, Y. Implementation of the blue moon ensemble method. *Chem-Bio Inf. J.* **2007**, *7*, 12–23.

(49) Flyvbjerg, H.; Petersen, H. G. Error estimates on averages of correlated data. *J. Chem. Phys.* **1989**, *91*, 461–466.

(50) Brüssel, M.; Di Dio, P. J.; Muñiz, K.; Kirchner, B. Comparison of Free Energy Surfaces Calculations from Ab Initio Molecular Dynamic Simulations at the Example of Two Transition Metal Catalyzed Reactions. *Int. J. Mol. Sci.* **2011**, *12*, 1389–1409.

(51) Spreafico, C.; Schiffmann, F.; VandeVondele, J. Structure and Mobility of Acetic Acid at the Anatase (101)/Acetonitrile Interface. *J. Phys. Chem. C* **2014**, *118*, 6251–6260.

(52) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc Natl Acad Sci USA* **2002**, *99*, 12562–12566.

(53) Laio, A.; Gervasio, F. L. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Rep. Prog. Phys.* **2008**, *71*, 126601.

(54) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008**, *100*, 020603.

(55) Barducci, A.; Bonomi, M.; Parrinello, M. Metadynamics. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1*, 826–843.

(56) Sutto, L.; Marsili, S.; Gervasio, F. L. New advances in metadynamics. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 771–779.

(57) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **2014**, *185*, 604–613.

(58) Raiteri, P.; Laio, A.; Gervasio, F. L.; Micheletti, C.; Parrinello, M. Efficient Reconstruction of Complex Free Energy Landscapes by Multiple Walkers Metadynamics. *J. Phys. Chem. B* **2006**, *110*, 3533–3539.

(59) Rohrdanz, M. A.; Zheng, W.; Clementi, C. Discovering Mountain Passes via Torchlight: Methods for the Definition of Reaction Coordinates and Pathways in Complex Macromolecular Reactions. *Annu. Rev. Phys. Chem.* **2013**, *64*, 295–316.

(60) Peters, B. Reaction Coordinates and Mechanistic Hypothesis Tests. *Annu. Rev. Phys. Chem.* **2016**, *67*, 669–690.

(61) Chen, W.; Tan, A. R.; Ferguson, A. L. Collective variable discovery and enhanced sampling using autoencoders: Innovations in network architecture and error function design. *J. Chem. Phys.* **2018**, *149*, 072312.

(62) Sultan, M. M.; Pande, V. S. Automated design of collective variables using supervised machine learning. *J. Chem. Phys.* **2018**, *149*, 094106.

(63) Kazaryan, A.; Baerends, E. J. Assessment of density functional methods for reaction energetics: Iridium-catalyzed water oxidation as case study. *J. Comput. Chem.* **2013**, *34*, 870–878.

(64) Nyhlén, J.; Duan, L.; Åkermark, B.; Sun, L.; Privalov, T. Evolution of O2 in a Seven-Coordinate RuIV Dimer Complex with a [HOHOH]- Bridge: A Computational Study. *Angw. Chem. Int. Ed.* **2010**, *49*, 1773–1777.

(65) Ertem, M. Z.; Gagliardi, L.; Cramer, C. J. Quantum chemical characterization of the mechanism of an iron-based water oxidation catalyst. *Chem. Sci.* **2012**, *3*, 1293–1299.

(66) Schilling, M.; Hodel, F.; Luber, S. Discovery of open cubane-core Structures for biomimetic $\{LnCo_3(OR)_4\}$ Water Oxidation Catalysts. *ChemSusChem* **2017**, *10*, 4561–4569.

(67) Sinha, V.; Govindarajan, N.; de Bruin, B.; Meijer, E. J. How Solvent Affects C-H Activation and Hydrogen Production Pathways in Homogeneous Ru-Catalyzed Methanol Dehydrogenation Reactions. *ACS Catal.* **2018**, *8*, 6908–6913.

(68) Ma, C.; Piccinin, S.; Fabris, S. Reaction Mechanisms of Water Splitting and H_2 Evolution by a Ru(II)-Pincer Complex Identified with Ab Initio Metadynamics Simulations. *ACS Catal.* **2012**, *2*, 1500–1506.

(69) Marcos-Alcalde, I.; Setoain, J.; Mendieta-Moreno, J. I.; Mendieta, J.; Gómez-Puertas, P. MEPSA: minimum energy pathway analysis for energy landscapes. *Bioinformatics* **2015**, *31*, 3853–3855.

(70) Dijkstra, E. W. A note on two problems in connexion with graphs. *Numer. Math* **1959**, *1*, 269–271.

(71) Park, J. M.; Laio, A.; Iannuzzi, M.; Parrinello, M. Dissociation Mechanism of Acetic Acid in Water. *J. Am. Chem. Soc.* **2006**, *128*, 11318–11319.

(72) Giorgino, T.; Laio, A.; Rodriguez, A. METAGUI 3: A graphical user interface for choosing the collective variables in molecular dynamics simulations. *Comput. Phys. Commun.* **2017**, *217*, 204–209.

(73) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. Electronic structure calculations on workstation computers: The program system turbomole. *Chem. Phys. Lett.* **1989**, *162*, 165–169.

(74) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988**, *38*, 3098–3100.

(75) Perdew, J. P. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev. B* **1986**, *33*, 8822–8824.

(76) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.

(77) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

(78) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785–789.